Comparing alternate approaches to calculating reliability for dichotomous data: The sample case of adolescent selection, optimization, and compensation

Napolitano, Christopher M ; Callina, Kristina S ; Mueller, Megan K

Abstract: Estimating reliability for scales or factors is an essential data analysis step in much of the research in developmental science. In this article, we demonstrate the importance of using the appropriate statistical method and underlying correlation matrix to estimate reliability for dichotomous data that represent a normally-distributed latent factor. We used an example case of three waves of adolescent data collected from responses to the Selection, Optimization, and Compensation questionnaire (SOC; Freund Baltes, 2002) of intentional self-regulation to illustrate how calculating composite reliability (or) using tetrachoric correlations provides a more accurate estimate of reliability when compared to both raw covariance-based , as well as raw covariance-based and tetrachoric correlation-based Cronbach's approaches. In addition, we describe methods for calculating each of these approaches to reliability estimation, and we offer suggestions for future researchers for estimating reliability for such dichotomous data.

DOI: https://doi.org/10.1080/10888691.2013.804372

Originally published at:
Napolitano, Christopher M; Callina, Kristina S; Mueller, Megan K (2013). Comparing alternate approaches to calculating reliability for dichotomous data: The sample case of adolescent selection, optimization, and compensation. Applied Developmental Science, 17(3):148-151.
DOI: https://doi.org/10.1080/10888691.2013.804372

Comparing Alternate Approaches to Calculating Reliability for Dichotomous Data: The Sample Case of Adolescent Selection, Optimization, and Compensation[1]

Christopher M. Napolitano

Kristina S. Callina

Megan K. Mueller

Tufts University

Submitted: 3/18/13

[1] This research was supported in part by grants from the National 4-H Council, the Thrive Foundation for Youth, and the John Templeton Foundation. The authors would like to thank the editor and reviewers for their helpful comments on a prior version of this article.

Address correspondence to: Christopher M. Napolitano, Institute for Applied Research in Youth Development, Tufts University, 301 Lincoln Filene Building, Medford, MA, 01255; Email: Christopher.Napolitano@tufts.edu

Abstract

Estimating reliability for scales or factors is an essential data analysis step in much of the research in developmental science. In this article, we demonstrate the importance of using the appropriate statistical method and underlying correlation matrix to estimate reliability for dichotomous data that represent a normally-distributed latent factor. We used an example case of three waves of adolescent data collected from responses to the Selection, Optimization, and Compensation questionnaire (SOC; Freund & Baltes, 2002) of intentional self-regulation to illustrate how calculating composite reliability (or É) using tetrachoric correlations provides a more accurate estimate of reliability when compared to both raw covariance-based É, as well as raw covariance-based and tetrachoric correlation-based Cronbach's ± approaches. In addition, we describe methods for calculating each of these approaches to reliability estimation, and we offer suggestions for future researchers for estimating reliability for such dichotomous data.

Calculating estimates of internal reliability for scales is an essential data analysis step in many life-span developmental research projects. Accurate estimates of internal reliability are critical, as misestimated reliability for a scale may result in incorrect interpretations of the relations of interest, or the unnecessary removal or replacement of items or scales from analyses. While some scholars have highlighted limitations to particular statistical techniques for computing reliability (e.g., Raykov, 1997; Sijtsma, 2009), accurately estimating reliability also depends on employing a suitable correlation matrix for assessing the data at hand (see also Geldhof, Preacher, & Zyphur, in press). In this article, we illustrate the importance of using the appropriate approach to estimating reliability for a particular type of data: dichotomous items that represent a normally-distributed item true scores.

This article is organized in four sections. We begin by describing and offering the strengths and limitations of four methods to calculate internal reliability. We next introduce the sample data for this paper, which are derived from dichotomous responses to the Selection, Optimization, and Compensation questionnaire (SOC; Freund & Baltes, 2002) of intentional self-regulation (ISR). Third, using three waves of data collected from adolescents, we calculate and compare the reliability estimates of SOC using four approaches to estimating reliability. We conclude by interpreting these results and by offering suggestions for estimating reliability for dichotomous questionnaire responses that represent a normally-distributed latent construct.

Cronbach's ± and KR-20

Easily computed in most statistical packages (e.g., SPSS, SAS), Cronbach's ± (1951) is the most common method for calculating reliability for continuous data. Cronbach's ± is calculated using the following formula:

$$\alpha = \frac{n^2 \bar{\sigma}_{ij}}{\sigma_x^2} \tag{1}$$

where α is a function of the average inter-item covariance within a scale ($\bar{\sigma}_{ij}$), the variance of the scale score (σ_x^2), and the number of items included in the scale (n). Based on the Kuder-Richardson Formula 20 (KR-20; Kuder & Richardson, 1937), Cronbach's ± is mathematically equivalent to KR-20 for scales with dichotomous items. While estimating reliability for dichotomous items using Cronbach's ± / KR-20 is convenient, at least two issues limit this technique's utility for estimating reliability for scales with dichotomous data that represent a normally-distributed latent construct.

First, the default setting for computing Cronbach's ± / KR-20 in popular statistical packages involves using raw convariances for computation, which assumes a binomial distribution for dichotomous data. For the type of data relevant to this research (dichotomous items that represent a normally-distributed item true scores), it is more appropriate to compute Cronbach's ± / KR-20 using the tetrachoric correlation matrix (Pearson, 1901), which assumes a normal distribution. While it is technically possible to compute Cronbach's ± / KR-20 using tetrachoric correlations with any statistical package that provides these values, this procedure can be cumbersome and is not common in the literature.

The second limitation of Cronbach's ± / KR-20 is based in the mathematical calculation of the values. Cronbach's ± / KR-20 assumes that all items are equally good representations of a scale's underlying latent construct (Novick & Lewis, 1967). This supposition, is termed *essential tau-equivalence*, and it is rarely tenable in the social sciences (Raykov, 1997). Because of these limitations, the typical approach to calculating Cronbach's ± / KR-20 using popular statistical packages will result in a lower-bound estimate of reliability.

Composite Reliability (É)

There are alternate methods of estimating reliability that involve using the true score variance of dichotomous items. A notable example is termed composite reliability, or coefficient É (McDonald, 1999), which can be calculated using the following formula:

$$\omega = \frac{(\sum \lambda)^2}{(\sum \lambda)^2 + \sum \theta} \quad (2)$$

where the numerator represents the scale score's true score variance and the denominator represents the total variance of the unit-weighted scale score. Composite reliability has been forwarded as superior to Cronbach's ± / KR-20; however, few applied examples involving the estimation of reliability using É exist in the life-span developmental literature.

Because composite reliability estimation is based on the true score variance of items, it is calculated using factor analysis. The default estimator for most factor analysis software is the maximum likelihood (ML) approach, which uses raw covariances to calculate parameter estimates. While using an ML-based estimator to calculate É provides accurate estimates of reliability for continuous data, it is more accurate to estimate the reliability of dichotomous data that represent a normally distributed true scores using a weighted-least squares-based (WLS) estimator and parameter estimates based on the tetrachoric correlation matrix (Rhemtulla et al., 2010). In sum, a WLS-based approach using tetrachoric correlations is the most appropriate method for estimating reliability for dichotomous data that represent normally distributed true scores. Computing É from tetrachoric correlations therefore provides a reliability estimate that is sensitive to heterogeneous item quality and uses values consistent with the categorical nature of the data.

The Selection, Optimization, and Compensation Questionnaire

To illustrate the utility of a WLS-based É approach to estimating reliability for dichotomous data that represent a normally-distributed latent construct, we will compare various reliability estimates for three waves of adolescent-reported data from the SOC questionnaire (Freund & Baltes, 2002) of ISR. We chose to use data from the SOC questionnaire for methodological and conceptual reasons. First, the SOC questionnaire format uses forced-choice dichotomous items that should each represent normally distributed item true scores. Prior research using this questionnaire has also reported low reliability, as calculated by raw-covariance-based Cronbach's ± (e.g., Mueller et al., 2011; Napolitano et al., 2011b; Schmid et al., 2011; Zimmerman, Phelps, & Lerner, 2007), despite the use of tetrachoric correlations for hypothesis testing.

Given the conceptual utility and empirical validity of the SOC construct, it is important to accurately assess the internal reliability of SOC in various samples across the breadth of the life span. The purpose of this research, therefore, is to illustrate that estimating reliability for dichotomous SOC data with a WLS-based É approach using tetrachoric correlations provides a more accurate estimate of reliability when compared ML-derived É, Cronbach's ± computed using raw covariances, and Cronbach's ± computed using tetrachoric correlations.

Method

Participants

We used data from the *4-H Study of Positive Youth Development*, a longitudinal study of youth and their families that, at this writing, includes 7,071 youth from 42 states. See Lerner et al. (2005) for full details of the methodology of the study.

The present research analyzes data from all youth who participated in at least two waves of data collection across Waves 6 to 8 (Grades 10 to 12). The Wave 6 (Grade 10) sample consisted of 694 participants, with a mean age of 15.8 years (*SD* = 1.00). The participants were 69.7% female, and self-reported as 79.1% White/Caucasian, 3.5% Hispanic/Latino, 4.0% African American, 2.2% Asian/Asian American, 0.6% Native American, 0.1% multiethnic/multiracial, and 10.5% inconsistently reported their ethnicities across waves of data collection.

The Wave 7 (Grade 11) sample consisted of 692 participants, with a mean age of 16.9 years (*SD* = .99). The participants were 68.6% female, and self-reported as 78.9% White/Caucasian, 3.5% Hispanic/Latino, 3.6% African American, 2.3% Asian/Asian American, 0.7% Native American, 0.1% multiethnic/multiracial, and 10.8% inconsistently reported their ethnicities across waves of data collection.

The Wave 8 (Grade 12) sample consisted of 701 participants, with a mean age of 17.9 years (*SD* = .93). The participants were 68.9% female, and self-reported as 79.0% White/Caucasian, 3.3% Hispanic/Latino, 4.0% African American, 2.3% Asian/Asian American, 0.6% Native American, 0.1% multiethnic/multiracial, and 10.7% inconsistently reported their ethnicities across waves of data collection.

Measure

We analyzed a nine-item composite to index SOC at each of the three waves of data collection. This nine-item composite includes a subset of items in the SOC questionnaire (Freund & Baltes, 2002). This composite was originally proposed by Gestsdóttir and Lerner (2007), and has commonly been used an index of ISR in the adolescent development literature (e.g., Bowers et al., 2011). The composite contains items from each of the three core SOC concepts: Selection, Optimization, and Compensation. Selection involves developing a goal hierarchy and the commitment to a goal or set of goals. Optimization refers to the acquisition and investment of means to achieve one’s goals. Compensation refers to the use of strategies that allow one to overcome goal-related obstacles in order to maintain a given level of functioning (Freund & Baltes, 2002).

Each item in the SOC questionnaire asks participants to indicate which of two people they are most similar, Person A or Person B. One person represents high levels of SOC, while the other indicates low levels of SOC. Thus, all items are scored dichotomously. A sample item is, *Person A: I concentrate all my energy on a few things OR Person B: I divide my energy among many things*.

Analyses

We calculated the reliability of the nine-item SOC composite across three waves of longitudinal data using four approaches: 1. Cronbach•s ± using raw covariances; 2. Cronbach•s ± using tetrachoric correlations; 3. É using raw coviarances; and 4. É using tetrachoric correlations. First, we calculated Cronbach•s ± using raw covariances using the “RELIABILITY” function in SPSS 21. Next, we calculated Cronbach•s ± using tetrachoric correlations in Mplus 7. To do so, we used the method described by Geldhof and colleagues (in press), which involved specifying three independent fully-saturated models that included only item covariances, with items specified as categorical and using the WLSMV estimator. We then input these values into the earlier-described Equation 1.

The final two approaches involved calculating É using estimates derived from two sets of confirmatory factor analyses (CFAs) in Mplus 7. The third approach to calculating reliability for SOC involved fitting three single latent factor CFA models of the nine-item SOC factor using the ML estimator, which utilizes raw covariances for estimation. The final approach to calculating reliability for SOC involved fitting three single latent factor CFA models of the nine-item SOC factor using the WLSMV estimator and specifying all items as categorical, which utilizes the tetrachoric correlation matrix for estimation.

Results

The results indicated that both approaches using raw covariances provided a lower reliability estimates for the 9-item SOC scale when compared to the methods that employed the tetrachoric correlations. This difference reflects the fact that the two sets of estimates made qualitatively different assumptions about the data and thus estimated qualitatively different parameters. The composite reliability (É) estimates computed using tetrachoric correlations demonstrated that the nine-item SOC composite was characterized by acceptable reliability in all analyses, and the tetrachoric-correlation-based Cronbach's ± estimates yielded similar values. In contrast, both raw covariance approaches provided much lower reliability estimates. Table 1 displays the reliability of the nine-item SOC scale as indexed by raw covariance-based Cronbach•s ±, tetrachoric correlation-based Cronbach's ±, raw covariance-based É, and tetrachoric correlation-based É for Waves 6, 7, and 8.

Insert Table 1 about here

Discussion

These results highlighted the utility of estimating reliability using tetrachoric correlations when items are dichotomous and represent normally distributed true scores. In contrast, Cronbach•s ± and É computed using raw covariances provided lower reliability estimates for the nine-item SOC factor across Waves 6, 7, and 8 because these data were not binomially distributed. While some recent research has estimated the reliability of SOC using tetrachoric correlations (e.g., Bowers,et al., 2011), raw-covariance-based reliability estimates for SOC have been used frequently in the developmental literature (e.g., Mueller et al., 2011; Napolitano et al., 2011b; Schmid et al., 2011; Zimmerman, Phelps, & Lerner, 2007) and may have resulted in potential misinterpretation of the scale’s psychometric characteristics.

The results also illustrated that Cronbach•s ± and É computed using tetrachoric correlations provided acceptable and relatively similar reliability estimates. We suggest that future research involving SOC, or other dichotomous items that each represent a normally-distributed true score, should calculate reliability using a tetrachoric correlation-based É approach for two reasons. First, computing É in factor analysis programs like Mplus is currently more efficient than computing Cronbach•s ± in these same programs. Second, as discussed earlier, while the estimates for these data were relatively similar, Cronbach•s ± generally provides a lower bound estimate of reliability by assuming tau-equivalence. Composite reliability based on tetrachoric correlations, on the other hand, is quickly calculated and provide the most accurate representation of the internal reliability of a construct of interest with dichotomous data.

References

Bowers, E. P., Gestsdóttir, S., Geldhof, G. J., Nikitin, J., von Eye, A., & Lerner, R. M. (2011). Developmental trajectories in adolescence: The role of parenting and implications for positive and problematic outcomes among diverse youth. *Journal of Adoelscence, 34* (6), 1193-1206.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16(3)*, 297–334.

Freund, A. M., & Baltes, P. B. (2002). Life-management strategies of selection, optimization and compensation: Measurement by self-report and construct validity. Journal of Personality and Social *Psychology, 82*, 642-662.

Geldhof, G. J., Little, T. D., & Colombo, J. (2010). Self-regulation across the life span. In M. E. Lamb, A. M. Freund, & R. M. Lerner (Eds.), *The handbook of life-span development: Vol 2. Social and emotional development* (pp. 116-157). Hoboken, NJ: John Wiley & Sons Inc.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (in press). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods.*

Gestsdóttir, S., Bowers, E., von Eye, A., Napolitano, C. M**.,** Lerner, R. M. (2010). Intentional self regulation in middle adolescence: The emerging role of loss-based selection in positive youth development. *Journal of Youth and Adolescence, 39*(7), 764-782.

Gestsdóttir, S., & Lerner, R. M. (2007). Intentional self-regulation and positive youth development in early adolescence: Findings from the 4-H Study of Positive Youth Development. *Developmental Psychology, 43,* 508-521.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*(3), 151–160.

Lerner, R. M., Lerner, J. V., Almerigi, J., & Theokas, C., Phelps, E., Gestsdóttir, S., et al. (2005). Positive youth development, participation in community youth development programs, and community contributions of fifth grade adolescents: Findings from the first wave of the 4-H Study of Positive Youth Development. *Journal of Early Adolescence, 25(1),* 17-71.

McClelland, M. M., Ponitz, C. C., Messersmith, E., & Tominey, S. (2010). Self-regulation: The integration of cognition and emotion. In R. Lerner (Series Ed.) & W. Overton (Vol. Ed.), Handbook of life-span development. Vol. 1: Cognition, biology and methods (pp. 509–553). Hoboken, NJ: Wiley & Sons.

McDonald, R.P. (1999). *Test theory: A unified treatment*. Hillsdale: Erlbaum.

Mueller, M. K., Phelps, E., Bowers, E. P., Agans, J. P., Urban, J. B., & Lerner, R. M. (2011). Youth development program participation and intentional self-regulation skills: Contextual and individual bases of pathways to positive youth development. *Journal of Adolescence, 34*(6), 1115-1125.

Napolitano, C. M., Bowers, E. P., Gestsdottir, S., Depping, M., von Eye, A., Chase, P., & Lerner, J. V. (2011). The role of parenting and goal selection in positive youth development: A person-centered approach. *Journal of Adolescence, 34* (6), 1137-1150.

Novick, M. R., & Lewis, C. L. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika, 32*, 1-13.

Pearson, K. (1901). Mathematical contributions to the theory of evolution: On the correlation of characters not qualitatively measurable. *Philosophical Transactions of the Royal Society of London*, 195, 1–47.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement, 21*(2), 173-184.

Rhemtulla, M., Brosseau-Liard, P. et Savalei, V. (2010). How many categories is enough to treat data as continuous? A comparison of robust continuous and categorical SEM estimation methods under a range of non-ideal situations. Retrieved from http://www2.psych.ubc.ca/~mijke/files/HowManyCategories.pdf.

Schmid, K. L., Phelps, E., Kiely, M., Napolitano, C. M., Boyd, M. J., & Lerner, R. M. (2011). The role of adolescents' hopeful futures in predicting positive and negative outcomes: Findings from the 4-H study of positive youth development. *The Journal of Positive Psychology, 6*(1), 45-56.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120.

Zimmerman, S., Phelps, E., & Lerner, R. M. (2007). Intentional self-regulation in early adolescence: Assessing the structure of selection, optimization, and compensations processes. *European Journal of Developmental Science, 1*(3), 272-299.

Table 1. *Reliability Estimates for 9-item SOC at Waves 6 to 8*

Wave	Raw covariance Cronbach's ±	Tetrachoric correlation Cronbach's ±	Raw covariance É	Tetrachoric correlation É
6	.620	.781	.641[*]	.793
7	.651	.792	.662	.802
8	.655	.802	.672[*]	.817

[*]The model fit for the Wave 6 and 8 models was poor (CFI = .865 and .888, respectively) in part because the ML estimator assumes a binomial distribution for these normally-distributed dichotomous data. Since the standard errors of a poorly-fitting model may be untrustworthy, these exact É values may somewhat biased. However, the Wave 7 model had acceptable fit (CFI= .93) and displayed a similarly lower-bound reliability estimates when compared to the same-Wave tetrachoric correlation É estimates.